

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-207685

(43)公開日 平成10年(1998)8月7日

(51) Int.Cl. <sup>®</sup>	識別記号
G 0 6 F 3/16	3 3 0
12/00	5 4 7
G 1 0 L 3/00	5 5 1

F 1		
G 0 6 F	3/16	3 3 0 Z
	12/00	5 4 7 H
G 1 0 L	3/00	5 5 1 Z

審査請求 未請求 請求項の数60 O.L (全 15 頁)

(21)出願番号 特願平9-333888  
(22)出願日 平成9年(1997)12月4日  
(31)優先権主張番号 08/759916  
(32)優先日 1996年12月4日  
(33)優先権主張国 米国(US)

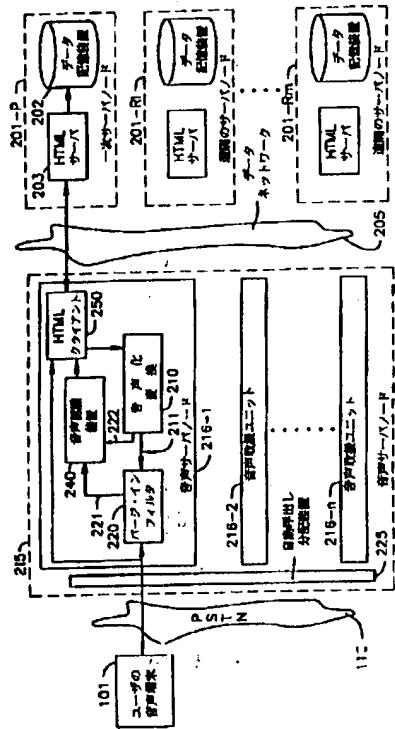
(71) 出願人 390035493  
エイ・ティ・アンド・ティ・コーポレーション  
AT&T CORP.  
アメリカ合衆国 10013-2412 ニューヨーク  
ニューヨーク アヴェニュー オブ  
ジ アメリカズ 32  
(72) 発明者 ダニエル ジットチャック メイヤー  
アメリカ合衆国 07059 ニュージャージ  
イ, ウォーレン, アップル ツリー レー  
ン 9  
(74) 代理人 弁理士 関部 正夫 (外3名)

(54) 【発明の名称】 ハイパーリンク化された情報との音声化されたインターフェースのためのシステムおよび方法

(57)【要約】

【課題】 本発明は、格納された情報とユーザーとが対話(Interaction)、特に、そのような対話が音声会話を介して行われる技術に関し、音声ノラウザの改善を提供することを目的とする。

【解決手段】 本発明は、音声の形式でユーザに対して提示される複数のハイパーリンクの中から1つのハイパーリンクの選択を容易にするものであり、該複数のハイパーリンクと1以上の他の語の音声発声を表す、ハイパーリンクの識別を含んだ第1の信号を該ユーザに提供し、該ユーザによって発声された音声を表す第2の信号の音声認識を実行する段階とからなり、該音声認識が、少なくとも2つの該ハイパーリンクに対応するエントリからなる認識装置の語彙を使うことによって実行されることを特徴とする。



# BEST AVAILABLE COPY

## 【特許請求の範囲】

【請求項1】 音声の形式でユーザに対して提示される複数のハイパーリンクの中から1つのハイパーリンクの選択を容易にする方法であって、該方法は、該複数のハイパーリンクと1以上の他の語の音声発声を表す第1の信号を該ユーザに提供する段階からなり、該信号は該ハイパーリンクの識別を含んでおり、該方法はさらに、

該ユーザによって発声された音声を表す第2の信号の音声認識を実行する段階とからなり、該音声認識が、少なくとも2つの該ハイパーリンクに対応するエントリーからなる認識装置の語彙を使うことによって実行されることを特徴とする方法。

【請求項2】 請求項1に記載の方法において、該信号を提供する段階は、テキストの集成に基づいて該第1の信号における該音声発声を合成する段階を含み、該集成が該複数のハイパーリンクを含むことを特徴とする方法。

【請求項3】 請求項2に記載の方法において、該テキストの集成がコンピュータ・ネットワーク・サーバによって提供されるドキュメントのテキストからなることを特徴とする方法。

【請求項4】 請求項3に記載の方法において、該ドキュメントがHTMLページからなることを特徴とする方法。

【請求項5】 請求項1に記載の方法において、該第1の信号で表される該音声発声の少なくとも一部分があらかじめ録音された人間の声からなることを特徴とする方法。

【請求項6】 請求項1に記載の方法において、該ユーザに提供されるハイパーリンクの該発声が、該他の語の該発声とは聴覚的に異なることを特徴とする方法。

【請求項7】 請求項1に記載の方法において、該第1および第2の信号が電話ネットワーク上で伝送され、そして少なくとも該信号のルーティングの部分が認識されたハイパーリンクに基づくことを特徴とする方法。

【請求項8】 請求項1に記載の方法において、該ハイパーリンクの該少なくとも2つが、該複数のハイパーリンクのすべてのハイパーリンクより少ないハイパーリンクを含むことを特徴とする方法。

【請求項9】 請求項1に記載の方法においてさらに、該ユーザに提供されるすべてのハイパーリンクのサブセットである該認識装置の語彙エントリーを選択する段階からなることを特徴とする方法。

【請求項10】 請求項1に記載の方法においてさらに、該複数のハイパーリンクのサブセットを規定する時間的なウィンドウに従って該ハイパーリンクの該少なくとも2つを選択する段階からなることを特徴とする方法。

【請求項11】 請求項10に記載の方法においてさら

に、該音声認識を実行する段階は、該ウィンドウ内で特定のハイパーリンクの時間的な位置に基づいて該ハイパーリンクの該少なくとも2つの中からの結果としての認識として特定のハイパーリンクを選択する段階からなることを特徴とする方法。

【請求項12】 請求項1に記載の方法において、該音声認識のための該認識装置の語彙からなる該複数のハイパーリンク・エントリーが、所定の時間間隔の間に該第1の信号の中に発生した該複数のハイパーリンクのサブセットに限定されることを特徴とする方法。

【請求項13】 請求項12に記載の方法において、該認識装置の語彙からなる該ハイパーリンク・エントリーの各々が、該第2の信号の未知の音声発声を表す確立が同じであることを特徴とする方法。

【請求項14】 請求項12に記載の方法において、該認識装置の語彙からなる該ハイパーリンク・エントリーの任意の1つが、該所定の時間間隔における該任意のハイパーリンクの時間的な位置に従って重み付けられている、該第2の信号の未知の音声発声を表す確立を有することを特徴とする方法。

【請求項15】 請求項1に記載の方法においてさらに、認識されたハイパーリンクに基づいて所定のアクションを実行させる段階からなることを特徴とする方法。

【請求項16】 請求項1に記載の方法においてさらに、認識されたハイパーリンクに基づいてトランザクションを初期化する段階からなることを特徴とする方法。

【請求項17】 請求項1に記載の方法においてさらに、認識されたハイパーリンクに基づいてトランザクションを実行する段階からなることを特徴とする方法。

【請求項18】 請求項2に記載の方法においてさらに、認識されたハイパーリンクに基づいてテキストの第2の集成を識別する段階からなることを特徴とする方法。

【請求項19】 請求項18に記載の方法において、該第2のテキストの集成がコンピュータ・ネットワーク・サーバ上にあるドキュメントのテキストを含むことを特徴とする方法。

【請求項20】 音声の形式でユーザに対して提示される複数のハイパーリンクの中から1つのハイパーリンクの選択を容易にするシステムであって、該システムが、該複数のハイパーリンクと1以上の他の語の音声発声を表す第1の信号を該ユーザに提供するインターフェースからなり、該信号は該ハイパーリンクの識別を含んでおり、該システムは更に、

該ユーザによって発声された音声を表す第2の信号の音声認識を実行する音声認識装置とからなり、該音声認識が、少なくとも2つの該ハイパーリンクに対応するエントリーからなる認識装置の語彙を使うことによって実行されることを特徴とするシステム。

【請求項21】 請求項20に記載のシステムにおい

3

て、該インターフェースは、テキストの集成に基づいて該第1の信号における該音声発声を合成するよう動作し、該集成は該複数のハイパーリンクを含むことを特徴とするシステム。

【請求項22】 請求項20に記載のシステムにおいて、該テキストの集成がコンピュータ・ネットワーク・サーバによって提供されるドキュメントのテキストからなることを特徴とするシステム。

【請求項23】 請求項20に記載のシステムにおいて、該第1の信号で表される該音声発声の少なくとも一部分があらかじめ録音された人間の声からなることを特徴とするシステム。

【請求項24】 請求項20に記載のシステムにおいて、該第1の信号の該ハイパーリンクの識別が、該他の語の該発声とは聴覚的に異なる形式で該ユーザに対して提供されるハイパーリンクの該発声を提供することによって実行されることを特徴とするシステム。

【請求項25】 請求項20のシステムにおいて、該第1および第2の信号が通信ネットワーク上で伝送され、そして少なくとも該信号のルーチングの部分が認識されたハイパーリンクに基づくことを特徴とするシステム。

【請求項26】 請求項20に記載のシステムにおいて、該認識手段に対する語彙のエントリーが、該ユーザに対して提供されたすべてのハイパーリンクのサブセットであるように選択されることを特徴とするシステム。

【請求項27】 請求項20に記載のシステムにおいてさらに、該複数のハイパーリンクのサブセットを規定する時間的なウインドウに従って該ハイパーリンクの該少なくとも2つを選択する手段を含むシステム。

【請求項28】 請求項27に記載のシステムにおいて、該音声認識装置は、該ウインドウ内で特定のハイパーリンクの時間的な位置に基づいて該ハイパーリンクの該少なくとも2つの中からの結果としての認識として特定のハイパーリンクを選択することを特徴とするシステム。

【請求項29】 請求項20に記載のシステムにおいて、該音声認識装置に対する該音声認識装置の語彙からなる該ハイパーリンク・エントリーが、所定の時間間隔の間に該第1の信号の中に発生した該複数のハイパーリンクのサブセットに限定されることを特徴とするシステム。

【請求項30】 請求項29に記載のシステムにおいて、該認識装置の語彙からなる該ハイパーリンク・エントリーの各々が、該第2の信号の未知の音声発声を表す確立が同じであることを特徴とするシステム。

【請求項31】 請求項29に記載のシステムにおいて、該認識装置の語彙からなる該ハイパーリンク・エントリーの任意の1つが、該所定の時間間隔における該任意のハイパーリンクの時間的な位置に従って重み付けられる、該第2の信号の未知の音声発声を表す確立を有す

10

20

30

40

50

4

ることを特徴とするシステム。

【請求項32】 請求項20に記載のシステムにおいて、該認識されたハイパーリンクに基づいて所定のアクションを実行させる手段を含むことを特徴とするシステム。

【請求項33】 請求項20に記載のシステムにおいてさらに、認識されたハイパーリンクに基づいてトランザクションを初期化するトランザクション初期化装置からなることを特徴とするシステム。

【請求項34】 請求項20に記載のシステムにおいてさらに、認識されたハイパーリンクに基づいてトランザクションを実行するトランザクション・プロセッサからなることを特徴とするシステム。

【請求項35】 請求項21に記載のシステムにおいて、認識されたハイパーリンクに基づいてテキストの第2の集成を識別する手段を含むことを特徴とするシステム。

【請求項36】 請求項35に記載のシステムにおいて、第2のテキストの集成がコンピュータ・ネットワーク・サーバ上にあるドキュメントのテキストからなることを特徴とするシステム。

【請求項37】 音声化された情報インターフェース・システムであって、該システムは、1つ以上の対応する情報リンクを有するテキストを含む情報のデータベースと、

該データベースと関連して情報が音声化された形式で提供されるよう動作する手段と、

該提供された情報に関連するユーザによる音声応答を認識する手段と、

該認識されたユーザ応答に応動して、該情報リンクのうちの少なくとも1つに関連する情報ヘシフトする手段とを含むシステム。

【請求項38】 請求項37に記載のシステムにおいて、該情報のデータベースが複数の情報層として配置され、そして該情報層間のリンクが該情報リンクによって提供されることを特徴とするシステム。

【請求項39】 請求項38に記載のシステムにおいて、該情報リンクが、任意の情報層の識別された情報セグメントとして提供されることを特徴とするシステム。

【請求項40】 請求項39に記載のシステムにおいて、該任意の情報層の中の該情報が複数のテキスト・語として提供されることを特徴とするシステム。

【請求項41】 請求項37に記載のシステムにおいて、該音声化された形式で情報が提供されるようする手段が、他の音声化された情報とは聴覚的に異なる方法で、該情報リンクが提供されるようする更なる手段を含むことを特徴とするシステム。

【請求項42】 請求項41に記載のシステムにおいて、該情報リンクが聴覚的に異なる方法で提供されるようする該更なる手段が、該他の音声化された情報とは

異なる反対の性の声によって該情報リンクが音声化されるように動作することを特徴とするシステム。

【請求項4 3】 請求項3 7 に記載のシステムにおいて、ユーザによる該音声応答が、該システムによってあらかじめ規定されたアクションに対する指令として構成されることを特徴とするシステム。

【請求項4 4】 請求項3 9 に記載のシステムにおいて、ユーザによる該音声応答が、該任意の層の該識別された情報セグメントの1 つの繰返しとして構成されることを特徴とするシステム。

【請求項4 5】 請求項3 7 に記載のシステムにおいて、該ユーザに対する該音声応答の確認を提供する更なる手段を含むことを特徴とするシステム。

【請求項4 6】 請求項3 7 に記載のシステムにおいて、該情報のデータベースにあるグラフィック情報をユーザに提供する更なる手段を含むことを特徴とするシステム。

【請求項4 7】 請求項4 6 に記載のシステムにおいて、該グラフィック情報が、グラフィック・アクセス・チャネル手段を経由して該ユーザに提供されることを特徴とするシステム。

【請求項4 8】 1 つ以上の対応する情報リンクを有するテキストを含む、格納された情報への音声化アクセスを提供する方法であって、該方法は、該情報の少なくとも一部分が音声化された形式で提供されるようにする段階と、

該提供された情報に関連するユーザからの音声応答を認識する段階と、該認識されたユーザ応答に応動して、該情報リンクのうちの少なくとも1 つに関連する情報へシフトする段階とからなることを特徴とする方法。

【請求項4 9】 請求項4 8 に記載の方法において、該情報が複数の情報層として配置され、該情報層間のリンクが該情報リンクによって提供されることを特徴とする方法。

【請求項5 0】 請求項4 9 に記載の方法において、該情報リンクが、任意の情報層の識別された情報セグメントとして提供されることを特徴とする方法。

【請求項5 1】 請求項5 0 に記載の方法において、該任意の情報層の中の該情報が複数のテキスト・語として提供されることを特徴とする方法。

【請求項5 2】 請求項4 8 に記載の方法において、該音声化された形式で情報が提供されるようにする該段階が、他の音声化された情報とは聴覚的に異なる方法で、該情報リンクが提供されるようにする副段階を含むことを特徴とする方法。

【請求項5 3】 請求項5 2 に記載の方法において、該情報リンクが聴覚的に異なる方法で提供されるようにする該サブ段階が、該他の音声化された情報とは異なる反対の性の声によって該情報リンクが音声化されるように

することを特徴とする方法。

【請求項5 4】 請求項4 8 に記載の方法において、ユーザによる該音声応答が所定のアクションに対する指令として構成されることを特徴とする方法。

【請求項5 5】 請求項5 0 に記載の方法において、ユーザによる該音声応答が、該任意の層の該識別された情報セグメントの1 つの繰返しとして構成されることを特徴とする方法。

【請求項5 6】 請求項4 8 に記載の方法において、該ユーザに対する該音声応答の確認を提供する更なる段階を含むことを特徴とする方法。

【請求項5 7】 請求項4 8 に記載の方法において、該格納された情報にあるグラフィック情報をユーザに提供する更なる段階を含むことを特徴とする方法。

【請求項5 8】 請求項5 7 に記載の方法において、該グラフィック情報が、グラフィック・アクセス・チャネル手段を経由して該ユーザに提供されることを特徴とする方法。

【請求項5 9】 情報の格納されたデータベースへのインターフェースを提供するシステムであって、該システムは、

一組のリンクされた情報層として、音声の型式で格納されている該情報のデータベースを提供する手段と、ユーザに該情報の特定の層を提供するよう動作する手段と、

該特定の層の情報に関連して該ユーザによる音声応答を認識する手段と、

該認識されたユーザ応答に基づいて該特定の層から1 つのリンクされた層へシフトするよう動作する手段とからなることを特徴とするシステム。

【請求項6 0】 情報の格納されたデータベースへインターフェースを提供するシステムであって、該システムは、

一組のリンクされた情報層として情報の該データベースを確立する手段からなり、そのような層間のリンクが特定の情報層のリンク語に関連しており、該システムは更に、

該格納された情報層と結合して動作し、任意の層の情報が音声化された形式で提供されるようにする手段からなり、該任意の層の該リンク・語が該任意の層の中他の情報とは聴覚的に異なる方法で提供されるものであり、該システムは更に、

該任意の層の該リンク・語の1 つに関連するユーザからの音声応答を認識する手段と、

該認識された音声によるユーザ応答に基づいて任意の層から該リンク・語へリンクされた別の層へシフトするよう動作する手段とからなることを特徴とするシステム。

【発明の詳細な説明】

【発明の属する技術分野】

【 0 0 0 1 】 本出願は、格納された情報とユーザとが対話(Interaction)、特に、そのような対話が音声会話を介して行われる技術に関する。

【 0 0 0 2 】

【 発明の背景】「 ブラウザ」として知られているソフトウェア・プログラムは、一般的に、ワールド・ワイド・ウェブ( WWW)として知られているインターネットの一部への簡単なアクセスを提供するために使われている。そのようなブラウザの例としては、ネットスケープ・コミュニケーション社( Netscape Communications Inc.)から入手できるネットスケープナビゲータ( Netscape Navigator)や、マイクロソフト社( Microsoft Corporation)から入手できるインターネット エクスプローラ( Internet Explorer)等がある。これらのブラウザは、コンピュータユーザがWWWからの情報を要求し表示することを支援する文字的・グラフィックス的なユーザ・インターフェースである。ブラウザによって表示される情報は、画像、文字、音声、グラフィックスおよびハイパーリンクを含むドキュメント( または「 ページ」)を含み、それはよく「 ハイパーテキスト( hypertext )」と称される。

【 0 0 0 3 】 ハイパーテキストは、情報を見いだせる別のアドレス( 通常は別のドキュメント)のテキスト形式のグラフィック表現である。そのような情報は普通、「 テキスト」によって伝えられる情報の内容と関連する。ハイパーテキストは、通常アドレスそのものではなく、ユーザの関心の対象となり得るいくつかの情報を伝えるテキストである。ユーザがハイパーテキストの一部分を選択( 例えば、マウスの「 クリック」によって選択)すると、ブラウザは通常、そのハイパーテキストに関連するアドレスに基づいてリバから別のドキュメントを要求する。この意味において、ハイパーテキストは、関連するアドレスにあるドキュメントへのリンクである。

【 0 0 0 4 】 従来のコンピュータ・ソフトウェア・ブラウザの他に、他の型式のブラウザが知られている。オーディオ・ブラウザ( Audio Browser)は、ユーザ( リスナー)に対し WWWのドキュメント・テキストを「 音読」することによってコンピュータ・ブラウザの機能に接近する。オーディオ・ブラウザは視力に障害のある人、あるいはコンピュータにアクセスできないが電話にアクセスできる人にとって特に有用である。テキストの音読は、従来のテキスト-音声変換( Text-To-Speech: TTS )技術によって、あるいは予め録音された音声を再生することによって行うことができる。ハイパーテキストが、その前後に「 ピープ音」などの可聴デリミタ( Audible Delimiter)によって、あるいはハイパーテキストをリスナーへ話すときの音声の特性の変化によって、リスナーに示される。リスナーがハイパーテキストに関係しているリンクされたアドレスへジャンプしたいときには、リ

スナーはDT MFトーン( すなわち、タッチトーン)によって応答するか、あるいは自動音声認識システムによって認識される「 ジャンプ」または「 リンク」などのコマンド語を話すことによって応答する。いずれの場合においても、オーディオ・ブラウザは、この応答を、リスナーに対して音読したハイパーリンクに関係するアドレスにあるドキュメントを検索するコマンドと解釈する。。

【 0 0 0 5 】

【 発明の概要】 本発明は、改善されたオーディオ・ブラウザに向けられている。従来のブラウザには、ハイパーリンクを選択するために単純なコマンド語またはトーンを使用しなければならない制限があることを本発明の発明者は認識している。特に、ハイパーテキストでリンクされたアドレスへジャンプする要望を示すために「 同じ」コマンドまたはトーンを用いるので、従来のオーディオ・ブラウザでは、そのリスナー( ユーザ)が「 次の」ハイパーリンク・テキストを示す「 前に」、任意のハイパーテキスト・リンクを選択しなければならない。ハイパーテキスト・リンクが連続的に速く現れることから、あるいは、追加のハイパーリンクを聞くまではどのハイパーリンクを選択するかをユーザが知ることができないことから、そのようなオーディオ・ブラウザのユーザは、巻き戻しおよび再生のコマンドを使って、ハイパーテキストの次の部分の音読の前に、読まれたが選択しなかったハイパーテキストを選択できるようにしなければならない。

【 0 0 0 6 】 本発明の発明者はさらに、「 目の見える」人に対するコンピュータ・ブラウザで採用されている音声認識技術の特徴が「 コンピュータの画面を見ることができない」人のためのブラウザを改善するのに有用であることを認識している。例えば、1995年6月5日に出願された米国特許出願第08/460,955号を参照されたい。

【 0 0 0 7 】 本発明の1つの実施例によると、例えばWWWドキュメントから得られる複数のハイパーテキスト・リンク( あるいは、もう少し詳細には「 ハイパーリンク語( Hyperlink Word)」)が、「 オーディオ」ブラウザに対する音声認識装置の語彙として用いられている。これらのハイパーリンク語を、そのオーディオ・ブラウザの「 話し声」の通常の課程でユーザに音読する。ユーザは、そのようなハイパーリンク語を、例えば「 話し声」に対する音声の特性の変化によって識別する。ユーザがハイパーリンク語を選択したい時、従来技術のオーディオ・ブラウザのようにコマンドを話すかDT MFトーンを使うのではなく、そのハイパーリンク語そのものを単に繰り返す。ドキュメントのハイパーリンク語のいくつか、またはすべてをその語彙として有している音声認識装置は、話されたハイパーリンク語を認識し、認識されたハイパーリンク語に関連してリンクされたアドレスへ

のジャンプを生じさせる。

【 0 0 0 8 】

【 発明の詳細な記述】現代の環境においては、ユーザとそのユーザに関心のある情報との間の電子媒体を経由したインターフェースは、ほぼ至る所にも存在するようになってきている。そのようなインターフェースの代表的なものが図1に示されている。そこでは音声端末( 例えれば、電話端末) 1 0 1 に位置するユーザは、代表的に公衆電話網( P S T N ) 1 1 0 として示されている通信経路を経由して、音声サーバノード ( Audio Serving Note ) 1 2 0 へのアクセスを得る。音声サーバノード 1 2 0 においては、音声サーバ1 2 2 が関連するデータベース( データ記憶装置1 2 1 ) に格納されている情報へのユーザに対するインターフェースを提供している。

【 0 0 0 9 】図に示されているように、ユーザは、データ端末( 例えれば、コンピュータ端末) 1 0 2 のような文字またはグラフィックを基礎とする媒体から所望の情報へのアクセスを得ることもできる。ユーザは、代表的に P S T N 1 1 0 として示されている通信経路を経由して、データサーバノード 1 3 0 へのアクセスを得る。データサーバノード 1 3 0 においては、データ・サーバ1 3 2 が関連するデータベース( データ記憶装置1 3 1 ) に格納されている情報へのユーザに対するインターフェースを提供している。

【 0 0 1 0 】そのような文字またはグラフィックを基礎とするインターフェース・デバイスから非常に複雑な多階層の情報源へのアクセスを提供することは既に知られているが、前述したとおり、従来技術で知られるような声を基礎とするインターフェースは、そのような情報の非常に限られた範囲だけにアクセスを提供することができる。

【 0 0 1 1 】しかし、複雑の程度及び詳細の程度を増加する( あるいは減少する) リンクされた層あるいは論理的な関係を指定しているネットワークのリンクのいずれかで構成されているテキストを基礎とする情報( トランザクション的なオプションを含む) を提供することは、この分野の技術においてよく知られている。情報が階層構造で構成されている場所において、そのような層の間のリンクが、特定の層に配置されているキー語またはフレーズに基づいて通常設定されており、このキー語の各々は通常、別の層にある関連する情報へのリンクを提供する。ここでの説明は、階層に格納されている情報へのアクセスに焦点を向けているが、この利用は例示的なものであり、本発明の範囲を限定することを意図するものではないことに留意されたい。事実、本発明は、論理的なリンクのすべての型式に適している。

【 0 0 1 2 】そのような階層的にリンクされた情報層のテキストを基礎とする集合を頻繁に用いている場合が、ハイパーテキスト・マークアップ・ランゲージ( HyperText Markup Language: H T M L ) として知られている方

法に見出される。H T M L は、ワールド・ワイド・ウェブに対する重要な機能を提供する。W W W によって、初期層、すなわち、「ホームページ」がユーザに提示され、そのホームページは、そのウェブ・サイトに関連する主題の事項やアプリケーションに関連する、比較的高いレベルの情報の記述を提供する。そのホームページ情報に関連するより詳しい情報または特定のトランザクションを追求したいユーザのために、そのホームページのテキスト中にキー語またはフレーズが強調表示されており、それらはより詳細な情報や特定のトランザクションにリンクされている。そのようなリンクがH T M L の機能によって提供されている。

【 0 0 1 3 】代表的なH T M L のアプリケーションにおいては、ユーザに対してテキストのページがパーソナル・コンピュータのモニタ表示され( 最初のそのようなページは通常はホームページと呼ばれている) 、そのテキスト中のハイパーテキスト( またはハイパリンク語) は特定の色又はアンダーライン付きで表示されるか、通常のテキストの活字とは異なる何らかの他の方法で表示される。そのようなハイパリンク語の下にある( すなわち、関連する) 情報にアクセスしたいユーザは、マウス・ポインタまたはカーソルでそのハイパーテキストを指定し、そしてマウスのボタンをクリックするか、あるいはキーボードの「 Enter 」キーを押すかのいずれかによって、下にある情報をアクセスする意志を知らせることなる。

【 0 0 1 4 】<1. 本発明に従った例示的な処理の紹介>1 つの実施例によると、相互にリンクされた情報の階層構造の集合に対する音声によるユーザ・インターフェースが、テキストを基礎とするリンクされたH T M L 層の集合( 以降ではH T M L 「 ページ 」 と呼ばれることがある) として、所望の情報データベースの初期確立を通じて提供される。これらのページは、単独のサーバに、あるいは複数のネットワーク化されたサーバに格納される。本発明の1 つの実施例によると、任意のH T M L ページのテキストが音声の形式に変換され、そのテキストのハイパリンク語は他のテキストとは区別される音声とされる。ユーザは、関連する追加の情報が必要であるポイントを表すハイパリンク語を繰り返す( すなわち、発声する) ことによって、この音声化された情報システムと対話( Interactive ) し、自動音声認識システムは、そのユーザによって与えられたハイパリンク語の発声を認識する。そのような与えられたハイパリンク語の認識に応動して、その与えられたハイパリンク語に対応する情報層へのジャンプが行われ、その後に、その新しい情報層のテキストが音声の形式に変換される。

【 0 0 1 5 】本発明のこの実施例によると、H T M L ページのテキストが音声形式に変換される。その音声化されたH T M L テキストが、この実施例における電話のリンクを含む種々の周知の通信リンクを経由して、ユーザ

に対して再生される。このようなテキストから音声形式への変換は非常によく知られており、通常はテキスト一音声合成器( Text-To-Speech synthesizer: TTS )によって実行されることになる。そのようなTTSシステムはそれ自身よく知られている。TTSシステムの例は米国特許第4,685,135号、第5,157,759号、及び第5,204,905号に開示されている。

【 0016 】実施例の音声化情報サービスとインターフェースするユーザは、追加の情報に関連するハイパーリンク語に向けられた応答によって、リンクされた情報の別の層を探索する興味を示すであろうから、音声化された情報は、ハイパーリンク語と他の音声化されたテキストとの間で聴覚的な区別をすることが望ましい。合成されたテキストの異なる部分についての異なる音声を生成する各種の方法が、TTS技術において知られている。本発明の一実施例を表すそのような方法の1つの例は、普通のテキストは男性の声で提供され、ハイパーリンク語は女性の声で提供されるものである。あるいはその逆にすることもできる。TTS技術における声の変更はよく知られた手法である。

【 0017 】ユーザは任意の層の情報についての音声化されたテキストを聴いており、そしてハイパーリンク語を聞いたときに、ユーザには2つの選択がある。ユーザは発音されたテキストを引き続き聴くことができる( HTML ページの連続する音読に対応している)。あるいは、そのハイパーリンク語がそのハイパーリンク語に関連する更に詳細な情報を必要とするかを入力要求( prompt )する場合、ユーザは、その語を繰り返すことによって、その語の選択を示すことができる。その音声によるユーザの応答は、通信リンクを経由して情報システムと関連する音声認識装置に伝えられる。その通信リンクはユーザに対して発音された情報を提供するのに用いられるのと同じ通信リンクであってもよい。そのような音声認識装置もこの分野の技術においてよく知られている。

【 0018 】本発明のシステムにおける音声認識装置の機能は、ユーザが発生した応答を、検討中の情報層のハイパーリンク語の1つ、またはシステムに何らかの動作をさせるように設定された少数の予約された「アクション」語( 例えば、コマンド )の1つのいずれかとして認識する機能である。従って、ハイパーリンク語はそのアクション語と一緒に、音声認識装置の語彙の一部として働く。アクション語は予約語であり、それ故にハイパーリンク語として使うことはできず、例えば、「停止( stop )」、「戻り( back )」、「開始( start )」、「より遅く( slower )」、「より速く( faster )」などであり、一般にシステム・オペレータによって設定される。もちろん、アクション語の集合は小さいことが好ましく、同じ集合がそのモデルのすべてのアプリケーションにわたって共通に維持されが望ましい。

【 0019 】本発明のシステムのための音声認識機能は

特に実装しやすい。というのは、この音声認識装置は、任意のある時点において少数の語彙、すなわち、ハイパーリンク語およびアクション語の語彙を認識するだけで済むからである。音声認識装置の性能を支援するため、ハイパーリンク語のスライディング・ウィンドウ( Sliding Window )を使って認識装置の語彙を定義することができ、任意のある時点で、その語彙は最近再生されたハイパーリンク語や以前に発音されたいくつかのハイパーリンク語を含むよう にすることができる( しかし、一般には、以前に再生されたリンク全ての合計より少ない )。従って、任意のハイパーリンク語と同じ時間間隔における追加の語( 追加のハイパーリンク語を含んでもよい )からなる音声認識装置の語彙に対してスライディング・ウィンドウ( これは発音者を追跡する )を使うことによって、語認識装置はその時間間隔の中に現われるハイパーリンク語( と、システムのアクション語 )を認識できればよい。さらに、それらのハイパーリンク語の発音を提供するTTSシステムは語認識装置と同じシステムの一部であるため、語認識装置およびTTSシステムはある種の音声データ、例えば、ハイパーリンク語の音素のシーケンスを共有することができ、それはTTSシステムと認識装置の「 ウィンドウ 」の同期化を保つのに役立つ。

【 0020 】ユーザによって話されたハイパーリンク語が語認識装置によって認識されることに応答して、特定のハイパーリンク語がユーザによって選択されたことを示す信号が生成される。純粋にテキストを基礎とするハイパーテキスト・システムで使われているものと同様の方法を使って、特定のハイパーリンク語のこの認識は、システムがそのハイパーリンク語にリンクされた情報層へジャンプするように働く。そのリンクされた層に到達すると、その層のテキストは、同様に、ユーザへの通信のために音声形式に変換され、その新しい層内のハイパーリンク語またはシステム・アクション語の選択に関する更なるユーザ応答の条件となる。ワールド・ワイド・ウェブのようなテキストを基礎とする既存の技術の場合と同様に、リンクされた情報層の1つまたはそれ以上が、他の場所において動作しているサーバに関連する記憶媒体に存在することができ、そのリンクが、第1のサーバとリンクされたサーバとの間の通信経路を経由して設定される。

【 0021 】また、任意の層に、情報の一部またはすべてを、ストリーム・オーディオ、例えばプログレッシブ・ネットワーク社( Progressive Networks, Inc. ) 製のリアルオーディオ( RealAudio™ )によってワールド・ワイド・ウェブ上で提供されるような予め録音された人間の声や格納された音声情報と することができる。この場合、ハイパーリンク語は、他のテキストに対して使われるものとは反対の性の声で、そのようなハイパーリンク語を録音することによって区別することができる。

【 0 0 2 2 】 < I I . 例のプロセスの実装> 図2において、本発明の方法を実行するシステムが示されている。この図を参照すると、対象となる情報データベースを表す一組のHT ML ページがデータ記憶装置2 0 2 の中に提供され、これは関連するHT ML サーバ2 0 3 とともに、一次サーバノード2 0 1 -P を構成する。しかしながら、情報セットの部分層または関連する部分を、リモートサーバノード2 0 1 -R 1 乃至2 0 1 -R m に格納することができ、そのような各リモートサーバノードは、HT ML サーバと関連するデータ記憶手段とをそれぞれ含んでいる。各リモートサーバノードは次に音声サーバノード2 1 5 ヘリンクし、データ・ネットワーク2 0 5 ( 例えは、インターネット ) を経由して他のサーバノードにリンクされる。

【 0 0 2 3 】 そのデータ・セットへのアクセスの要求に応答して( 例えは、ユーザの音声端末1 0 1 からP S T N 1 1 0 を介しての電話呼出しの到着によって ) 、音声サーバノード2 1 5 の自動呼出し分配装置2 2 5 は、利用できる音声取扱ユニット2 1 6 -1 をサービス要求に割り当てる。割り当られた音声取扱ユニットにおいて、HT ML クライアント2 5 0 は、一次サーバノード2 0 1 から呼び出されるべきHT ML ページの最初のページ( 「 ホームページ 」 ) が、割り当られた音声取扱ユニットによってさらに処理されるようにする( 一次サーバノード2 0 1 は音声取扱ノード2 1 5 と一緒に配置してもよい。 ) 。HT ML のホームページ( これは、一次サーバノード2 0 1 にあるデータ記憶装置2 0 2 から音声取扱ユニット2 1 6 -1 にあるHT ML クライアント2 5 0 へHT ML サーバ2 0 3 によって供給される ) は通常、次に音声化変換手段2 1 0 によって音声形式に変換される。通常これはT T S システムによって実現される。いくつかの、あるいはすべてのHT ML 「 ページ 」 の音声化された形式が、必ずしもユーザアクセス、ユーザ要求の直後である必要はないが、そのアクセス、要求の前に得られていて格納されていてもよいことに留意されたい。この分野の技術においてよく知られているキャッシング技術は、どの音声形式が予め格納されているか、ユーザ要求に応答してどれが発生されたかを決定することができる。

【 0 0 2 4 】 HT ML ホームページから音声化されたテキストは、次に通信リンク2 1 1 を介してバージ・イン・フィルタ( Barge-In Filter ) 2 3 0 へ送信され、そこからユーザの音声端末1 0 1 を経由してそのユーザが聞くことができる。ユーザが音声化変換手段によって発音されているHT ML ページを聴いている時、そのユーザは、追加のあるいは関連する詳細を得たい( あるいは以下に説明するようにトランザクションをトリガしたい ) と思うハイパーリンク語を聞くことができる。そのような追加のあるいは関連する詳細に対する要求を示すべく、ユーザは、ユーザの音声端末1 0 1 を介してハ

イパーリンク語を繰返す( 話す ) 。ユーザからのその音声応答は、バージ・イン・フィルタ2 2 0 を通じて処理され、通信リンク2 2 1 を介して音声認識装置2 4 0 へ送信される。

【 0 0 2 5 】 バージ・イン・フィルタ2 2 0 の重要な機能は、ユーザによって発声された語( 音声化変換手段によって発音された語を除く ) だけが、音声認識装置2 4 0 に入力されることを保証することである。そのようなバージ・イン・フィルタは、この技術分野において知られており、そして既知の源( 音声化変換手段 ) から生成された電気信号を、その既知の源とユーザが発した語とを混合した全てから差し引くことによって動作する。この開示の目的のために、バージ・イン・フィルタもまた、ユーザとその音声取扱ユニットとの間の伝送経路にある不完全性を補償するエコー・キャンセラとして働くことを理解されたい。

【 0 0 2 6 】 音声認識装置2 4 0 は、音声化変換手段2 1 0 から通信リンク2 2 2 を介してその認識語彙( ユーザが繰返し発声するであろうハイパーリンク語を伴う ) を同期化する。音声認識装置によって選択されたハイパーリンク語の認識に応答して、その語に関連する信号が、その認識装置からHT ML クライアント2 5 0 へ送信され、HT ML クライアント2 5 0 はその信号をHT ML サーバに対する適切なコードに変換し、そのコードはその選択されたハイパーリンク語へリンクされた情報層及び位置に対してハイパーリンクが設定されるべきである示す。この動作は、そのハイパーリンク語を指しているカーソルをマウスでユーザがクリックし、システムがそれに対して応答するのに似ている。

【 0 0 2 7 】 図3は、図2で示されてた特徴のある機能のいくつかの詳細を示している。特に、図3は、従来のテキスト-音素変換プロセッサ3 1 5 と音素-音声変換プロセッサ3 1 7 を含む音声化変換2 1 0 、HT ML のドキュメント・ページから利用できるテキストのストリームについて動作するハイパーテキスト識別プロセッサ3 1 0 、識別されたハイパーリンクと音素のストリングとの相関のためのハイパーテキスト-音素相関器3 2 0 、及びハイパーテキスト・テキストの識別されたシーケンスのどれがその音声認識システムに対する語彙の一部として音声認識プロセッサ3 5 0 によって使われるべきかを決定するウインドウ・フィルタ3 3 0 のT T S 処理を行う機能を示している。

【 0 0 2 8 】 記述されている実施例に従って、任意のHT ML ドキュメント・ページ( これはシステム・ユーザに対する音声での提示のためのものである。 ) が、HT ML クライアント2 5 0 によって一次サーバノード2 0 1 から呼び出され、それ以降の処理のために利用できるようになる。与えられたHT ML ドキュメントのページがハイパーテキスト識別プロセッサ3 1 0 によって解析され、そのページのハイパーテキストが識別される。

ハイパーテキスト識別プロセッサ310からの出力が、ハイパーテキスト-音素相関器320へ提供され、そしてハイパーリンク語とHTMLページの他のテキストとの間の音声の区別を容易にするために、その出力から得られる信号が音素-音声変換プロセッサ317へ提供される。

【0029】ドキュメント・ページ上のテキストが、音声の変換のために音声化変換器( TTS )システム210に対しても提供される。これに、テキスト-音素変換プロセッサ315によるテキストを音素のシーケンスに変換する段階と、音素-音声変換プロセッサ317による音素を音声へ変換する段階の、通常の2つの段階によって行われる。

【0030】相関付けられたハイパーテキストおよび音素のシーケンスがウィンドウ・フィルタ320に提示され、ウィンドウ・フィルタ320は、任意の時刻までにユーザーに対して再生されたどのハイパーリンク語/フレーズが音声認識装置の語彙(システムのアクション語と共に)を形成するかを識別する。このウィンドウ・フィルタ330は、過去のある期間(例えば、秒数または語数で測定することができる)における最近再生されたハイパーテキストとそれ以前の全てのハイパーテキストを選択する。ウィンドウ・フィルタ330は、ユーザーに最近再生された語に関する同期情報を、通信リンク318を経由して音素-音声変換プロセッサ317から受信する。そのウィンドウ・フィルタ・プロセスの結果、即ち任意のウィンドウの持続期間内で発生したハイパーリンク語又はフレーズのシーケンスが、そのような音声の音素モデル(通常は別個に訓練された隠れマルコフモデル( Hidden Markov Models )として実行される)とともにデータベース340に格納される。当然、データベース340は、システム・アクション語の音素モデルも含んでいる。従来の自動音声認識プロセッサ350は、ユーザーから(ページ・イン・フィルタ220および通信リンク221を経由して)未知の音声を受信し、ハイパーリンク語またはシステム・アクション語の現在の語彙の1つとしてその音声を認識するように動作する。音声認識プロセッサ350はデータベース340と対話して、従来、例えばビタビ( Viterbi )スコアリングをそのデータベースの中の各種のモデルによってその未知の音声に対して行う。ハイパーリンク語又はフレーズ、又はシステム・アクション語の認識に応答して、選択されたハイパーリンク語(例えば、それと同等のHTML「ページ」の検索)またはシステム・アクション語に対する適切な動作のために、その認識システムの出力が一次サーノード201に提供される。

【0031】ウィンドウ・フィルタ330は、所定の時間ウィンドウで発音されたすべてのハイパーリンク語を等しい確率でその音声認識装置の語彙の中に受け入れる、平坦な重み付けとすることができます。代わりに、ウ

ィンドウ・フィルタは、最近発音されたハイパーリンク語を、以前に認識ウィンドウで発音された語より高い確率でその音声認識装置の語彙の中に受け入れる、時間で定義された「コンテキスト的なスムージング」( contextual smoothing )を提供することもできる。これらの確率は、認識を実行する時に音声認識装置350によって考慮される。

【0032】ある種のシステム・アクション語は、音素-音声変換手段の動作(例えば、「より速く」、「より遅く」、...)を参照する。そのような語が音声認識プロセッサ350によって認識されると、それぞれを識別する信号が適切な動作のために音声化変換手段へ送信される。

【0033】ユーザーが経験する範囲内の任意の時点において、テキストから音声への変換音声ではなく、予め録音された音声または音声コンテンツ(例えば、音楽)が使えることも理解すべきである。発音されるテキストではなく人間の声が望ましいとき、そのシステムの動作が図4に示されている。図の中で見出せるように、この実施例におけるデータ源は、ストリーム音声サーバ410とともにHTMLサーバ201から構成されている(各サーバは適切な記憶手段を含んでいる)。HTMLサーバ201とストリーム音声サーバ410は、単独のサーバあるいは別々のサーバとして実行でき、そしてそれが複数の物理的サーバから構成されており、それらが同じ場所にあっても、あるいは離れた場所にあってもよいことに留意されたい。HTMLサーバ201によって提供されるデータは、以前に記述された実施例の場合と同様に、テキストのHTMLページである。しかし、ストリーム音声サーバに対し、データの内容は、そのユーザーが利用できるハイパーテキスト・データの集合の全て、あるいはその一部分に対応する予め録音された音声セグメントからなる。そのような音声セグメントは通常、正確な音読スクリプト( reading script )として、ハイパーテキスト・データの素材を人間が録音することによって設定される。1つの実施例においては、問題のデータのテキスト部分が男性の声で読まれて(記録され)、そしてハイパーリンク語は女性の声(男性の声とは明確に異なるピッチを有する)で読まれる。リンクを設定できる全てのセグメントは別々に録音される。ストリーム音声セグメントの再生はHTMLサーバによって制御される。

【0034】この実施例に対するシステムの動作は、図3に対して記載されているように進行する。ただし、ユーザーには発音された音声の代わりに、(少なくとも選択されたデータ・セグメントに対して)ストリーム音声の再生が提示される。通信リンク310上で再生されたすべてのハイパーリンク語が、ハイパーリンク音声識別装置417を通じてハイパーリンク語・テキストおよび音声同期手段420に対して与えられる。ハイパーリンク

音声識別はハイパーリンク語の音声を他のテキストに対する音声から識別するように動作する。この実施例においては、女性の声(ハイパーリンク語)を男性の声(他のテキスト)から識別する。前に述べたように、ハイパーリンク・テキスト識別手段310は、この場合はハイパーリンク語・テキストおよび音声同期化手段420を通してハイパーリンク語(テキスト形式)を供給し、それはこの分野の技術においてよく知られている方法で、同じ語のテキスト・バージョンと一緒にストリームオーディオハイパーリンク語の進行を追跡し、それによってウィンドウ・フィルタ330に対して必要な同期信号を提供する。ユーザは同じ方法でそのシステムとインターフェースし、そして音声認識手段は前と同じように動作する。ハイパーリンク語が認識されると、HTMLクライアントは前と同様にトリガされ、HTMLサーバはストリーム音声サーバがその要求されたあらかじめ録音されているセグメントへ移動するようにさせ、そしてその新しいセグメントをユーザに対して再生し続けるようになる。

【0035】<III. 発明の方法の応用>本発明の実施例は、従来の音声情報システムに付随する多くの問題を解決することができる。例えば、従来の音声情報システムは設計および使用が難しいことが多い。この困難性のために、リスナーがタッチ・トーン・キーを押すことによって選択しなければならないような、よくあるネストされた階層形式において、リスナに対して各種のオプションを提示するための「ユーザ・フレンドリ」なシステムを設計する問題が抑制される。この設計の仕事の困難性は、それ自身、例えば銀行または仲買い業の機関における自動化されたトランザクション・システムを初めて使うようなユーザが遭遇する。ユーザは音声「メニュー」の入れ込まれた階層は、ナビゲートするのが難しいことに不満を洩らす場合が多い。それと対照的に、本発明は情報をナビゲートして所望のオプションを選択するためのもっと直観的なインターフェースを提供する。本発明によると、ユーザは自分が必要とするオプションをしゃべり、システムとより直観的に(例えば、手を使わなくてよい、目を使わなくてよい)正しく対処できるようになる。さらに、本発明の方法によると、ユーザは特定のオプションを選択する時に利用できるオプションに気付くことがずっと多くなる。というのは、任意の点において利用できる、提示される複数の話し言葉のオプションが利用できるからである。多くの従来技術の方法におけるように、概念を番号と関係付ける必要はない。

【0036】本発明は自由様式、制約のないフレーズの認識に関する現代水準の音声認識システムの問題を解決することもできる。繰り返されるべき話し言葉のハイパーリンク語に上ってブラウザを表示することにより、そのシステムはハイパーテキストの選択においてリスナーによって話される可能性の高い語の限定された集合を前

もって「知る」。従って、このシステムは音声情報システムの設計者が工夫する可能性のある実質的どんな話された語やフレーズも認識することができる。設計者は認識装置の精度を維持するだけのために認識、あるいはいくつかの代替方法の中で使うために少数の語彙(その音声情報システムのコンテキストに対応している)を選択することに制限されない。

【0037】また、ウィンドウ・フィルタを採用している実施例では、認識装置の語彙を一時的に制限して最小化することによって、高度な音声認識性能を容易に得ることができる。従って、その認識装置は常にすべての語(設計者によってハイパーテキストとして選択された)を認識しようと試みることはない(あるいはその必要がない)。これによって認識装置の性能が改善される。というのは、認識装置が長時間にわたってそれぞれの可能な同義語を認識する必要があるすべての語が存在するために語彙が大規模になる時、正しい認識はますます困難になるからである。

【0038】また、本発明によって、音声情報システムの設計者は豊富なHTMLオーサリング・ツールを利用することができる。

【0039】本発明を実施することの他の利点としては、拡張されたHTML媒体において後で複製することができる情報空間を通じての記録された経路の指定がある。例えば、ユーザは電話を使って情報空間の中をナビゲートすることができ、その後、テキスト「および」関連付けられた画像(同じ経路を辿って得られる)をFAXマシンに対して、あるいは電子メールのメッセージに対する付属物として配達するようシステムに指令することができる。電話機のユーザおよび目が不自由なPCのユーザに対してWWWのペーパーをオープンにすること、音声メッセージングおよび電子メールの統合化された使用、およびタッチトーン電話機を採用しない場所にある世界において音声情報システムのより一般的な適用を提供する。

【0040】<IV. 結論>記憶された情報によって音声による対話をを行うためのシステムおよび方法が説明されてきた。その中で、実質的により簡単で、より直感的なユーザ・インターフェース以外に、従来技術によって処理されるよりずっと複雑な情報の集合の表示を提供する。本発明の1つの応用例においては、ユーザまたは潜在的なユーザの集合に対して利用できる情報の収集を行いたいエンティティによって、その情報が一組のリンクされたHTMLページの中に作り込まれるようにすることができる。そのHTMLデータは1つまたはそれ以上の取扱ノードに関連する記憶媒体の中にロードされることになる。無料の電話番号などの取扱ノードにアクセスするための手段が確立される。通常、情報の集合の利用可能性に関する情報(アクセスのための手段以外に)が公開され、ユーザ又は潜在的なユーザに対して公示され

る。取扱ノードにアクセスする時、ユーザはHTMLのデータベースの「ホームページ」の中に現われるテキストが発音されることによって迎えられる。この場合、そのホームページの中のハイパーリンク語は通常のテキストの発音方法とは明確に異なる方法で発音される。次にユーザはさらに詳しい情報が求められる（そのハイパーリンク語が発音された後、調整可能な時間のウィンドウの間に）ようなハイパーリンク語を聞いた後、そのハイパーリンク語を繰り返すことによって「バージ・イン (b  
arge in)」する（入り込む）。そのハイパーリンク語の「バージ・イン」の繰返しが、その取扱ノードに関連付けられている音声認識装置によって認識され（その時間ウィンドウの中の「アクティブ (active)」などの複数の語から）、そしてその特定のハイパーリンク語の選択を示している信号に変換され、サーバがそのハイパーリンク語に対してリンクされているHTMLポイントに対してハイパーリンクを生成するようにさせ、あるいは株の売買などのトランザクションをトリガするため、あるいはユーザの電話をそれ以降の会話のために別の電話にリンクするようになる。

【0041】本発明の実施例が詳細に説明されてきたが、追加の請求項によって定義されているような本発明の精神および範囲から逸脱することなしに、各種の変更および置き換えが可能であることを理解されたい。特に、このシステムはユーザによって発音されたハイパーリンク語を認識した時に、その語がそのユーザの選択の確認としてユーザに対して繰り返して送り返されるように変更することができる。「間違い (wrong)」または「停止 (stop)」など、短時間の間のユーザの応答が存在しない時に、このシステムはその語に対してリンクされるHTML層に対するハイパーリンクを実装するように進行する。ここに記載されているシステムおよび方法を追加の変更として、グラフィック・データを含んでいるHTMLページ（それはもちろん、口頭では伝えることができない）は、そのような画像の存在を示すために「画像がここにある (image here)」などのフレーズが発音されるように構造化することができる。追加の機能として、このシステムはそのような画像に关心のあることを示しているユーザに問合させてそのユーザのFAX番号を提供せんようにし、その問題の画像を含んでいるページのFAXによるコピーがそのユーザのFAX装置に送信されるようにすることができる。さらに変更として、そのデータの部分を音声の形式で記憶し、そしてその音声データの提示がそのユーザに対して行われるよう

にし、HTMLクライアントに対してリアルタイムのデジタル化されたオーディオ情報を提供するため、ストリーム音声と呼ばれるよく知られているWWW技法によってサービング・ノードに対する接続を設定する。

【0042】さらに、ユーザが本発明に従って音声化されたHTMLホームページから下位の音声化された詳細情報、あるいは関連付けられた層にナビゲートするプロセスは、例えば、自動化された呼出し案内システム付きの対話型の音声応答（Interactive Voice Response: I VR）システムの利点のすべてを提供するが、IVRのメニュー構造を制限したり無効にしたりすることなしに行われる。代わりに、そのナビゲーションのプロセスは特定の内容の如何にかかわらず、ユーザがホームページから関心のある情報を表示する層に入り込む、ワールド・ワイド・ウェブのテキストをベースとするナビゲーションの配置と本質的に同じ方法で、首尾一貫して動作する。そして、よく知られているように、そのWWWのHTMLシステムは高度に多様な情報アクセス媒体を提供するだけでなく、実質的にユーザ・フレンドリーである本質的に直感的なナビゲーション方式を備えていることも示されている。従って、ユーザはこのインターフェースに対するモデルを一度学習するだけで済み、その後、そのモデルを使っている別のデータベースとの対話が、そのデータベースの内容とは無関係に、対応しているナビゲーション方式を提供することが分かる。さらに、ハイパーリンク語にアクセスすることによって到達される下に隠れているデータは、学習したモデルの場合と対応している方法で行動する。

【0043】HTMLページに关心のある情報をオーディオリングする際、認識システムの性能を高めるため、および本発明の方法をユーザにとってより有用であるようにするために、そのハイパーリンクの語やフレーズは比較的「コンパクト」（すなわち、普通は1つまたは2つの語を含んでいる）であって数が少ないことが好ましい。

#### 【図面の簡単な説明】

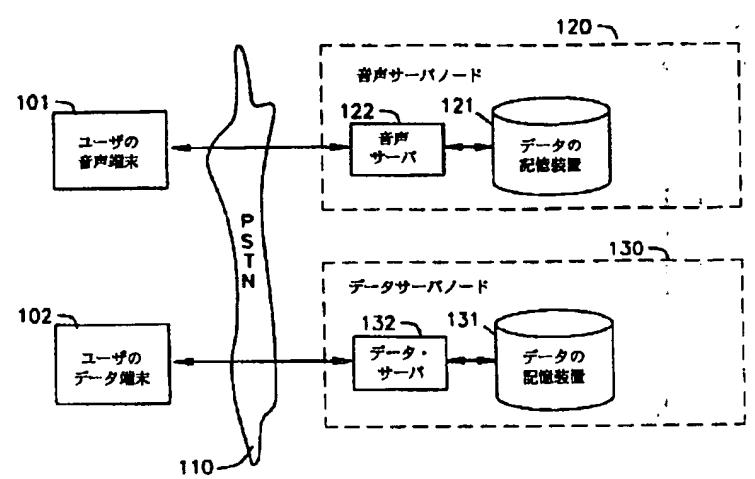
【図1】従来技術の情報アクセス・システムを示す図である。

【図2】本発明の音声による情報アクセスを示す図である。

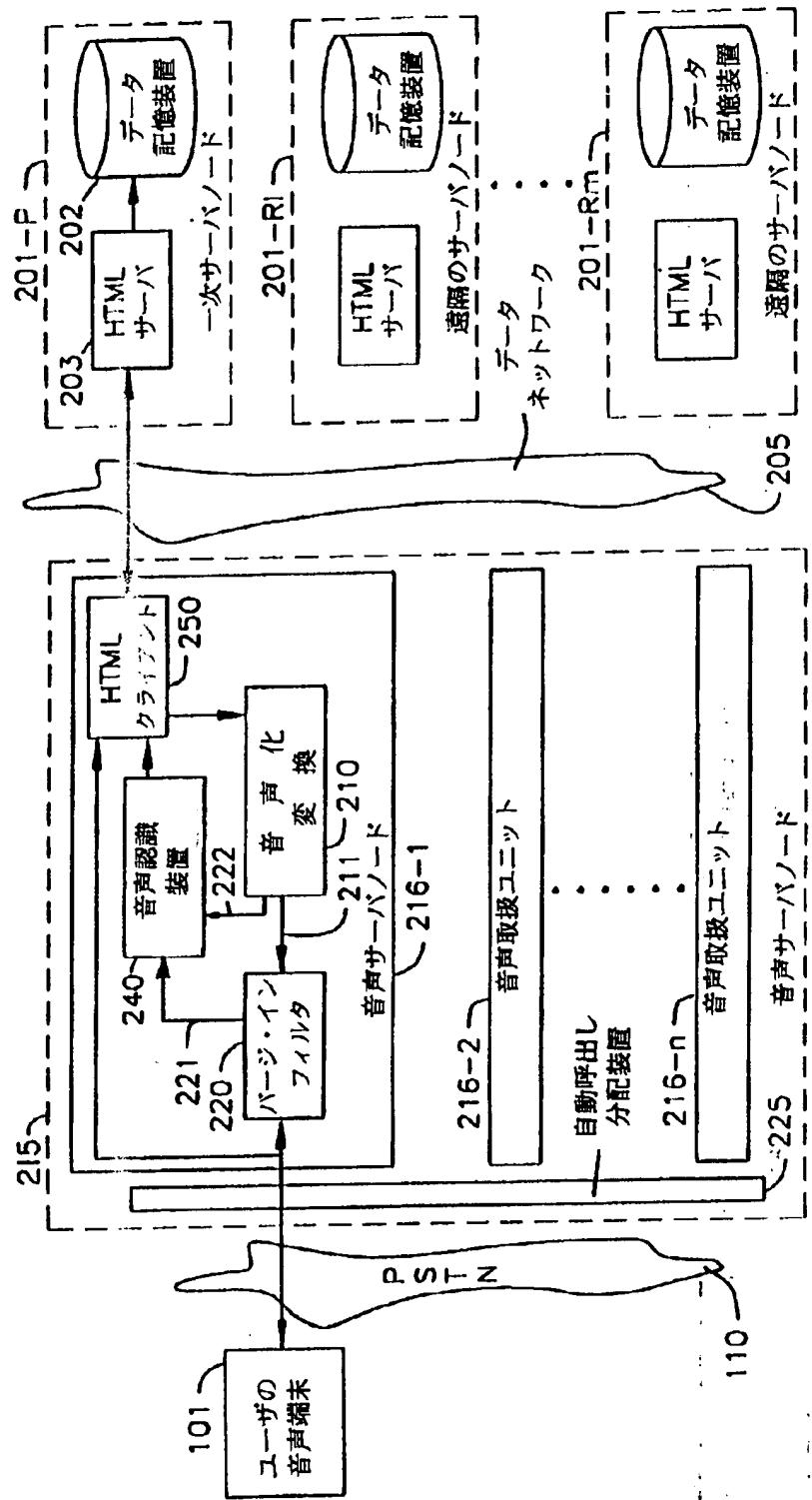
【図3】図2に略図が示されている機能のいくつかの異なる詳細を示す図である。

【図4】あらかじめ録音された音声または他の音声内容として情報が提供される場合の本発明のシステムの一実施例を示す図である。

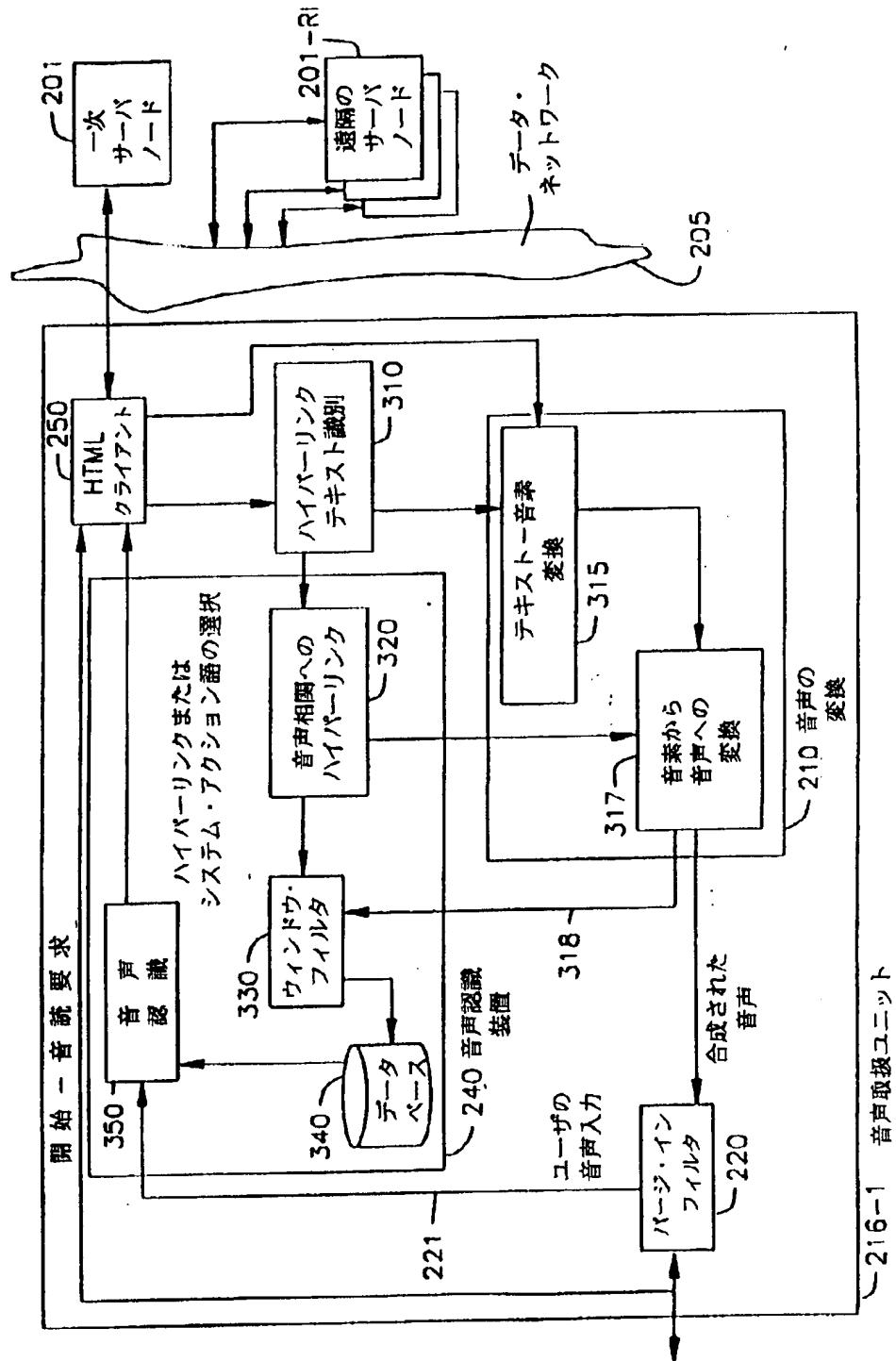
【 図1 】



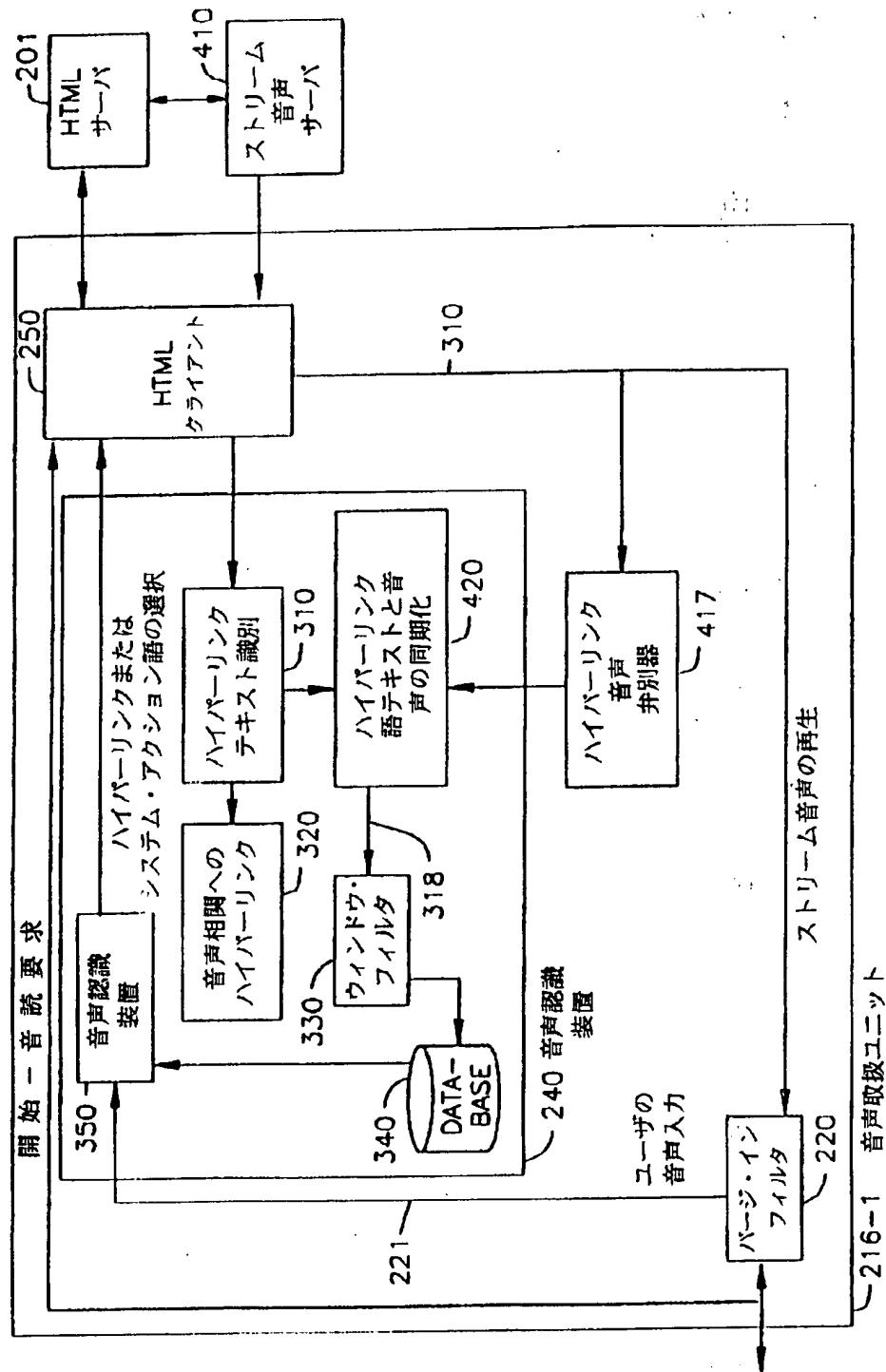
【図2】



【図3】



【図4】



-216-1 音取扱ユニット

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image  
problems checked, please do not report these problems to  
the IFW Image Problem Mailbox.**